DOCUMENT RESUME

ED 151 332 SP 012 323

AUTHOR Hsu, Yi-Ming; White, William F.
TITLE Interactions Between Teaching Performance and Student Achievement.
PUB DATE Mar 78
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Toronto, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Rating; *Behavior Rating Scales; *College Faculty; Effective Teaching; Higher Education; Interaction Process Analysis; *Student Evaluation of Teacher Performance; Student Teacher Relationship; Teacher Evaluation

ABSTRACT

There are two purposes for this study: first, to examine the relationship between college students' achievement and their ratings of instructors; second, to validate the two selected evaluation instruments that were designed specially for assessing teaching performance at the higher education level. Two evaluation inventories were selected for this study. One was the Inventory of Student Perceptions of Instruction (ISPI); the other, the Instructional Improvement Questionnaire (IIQ). The ISPI was given about half way through the fall semester, while the IIQ was distributed toward the end of the semester to the same classes. While it is clear that personal warmth in a teacher carries a rather significant weight in determining the rating behaviors of students and that higher ratings positively correlate with higher grades, these two evaluation methods counterbalanced these factors. At the end of the semester, three separate achievement scores were obtained from each instructor for all the students participating in the study. An analysis of the results indicated that college students tend to rate their professors more favorably if they felt they had learned a great deal and achieved higher grades in the course. They do not evaluate their professors as teaching in isolation, irrelevant of their own performance. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED151332

Interactions Between Teaching Performance and Student Achievement

Yi-Ming Hsu William F. White

West Chester State College

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Yi-Ming Hsu

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND THE ERIC SYSTEM CONTRACTORS."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada
March 1978

SP012 323

Interactions Between

Teaching Performance and Student Achievement

The use of student ratings as a primary means for evaluating classroom instruction has been believed to be pertinent to the measures of teaching effectiveness. It has served as the traditional need for providing a major data base on which such important administrative and personnel decisions as salary increments, ranks in promotion, and even considerations in granting tenure were to be made. As a consequence, the practices of the unended student evaluations have stirred up rather sensitive and vital concern to nearly everyone in all levels of school systems.

This concern is much more obvious in colleges and universities in which numerous evaluation instruments or rating scales have been designed and developed specifically for the purpose of appraising the quality of the instructional programs delivered in classrooms (Benton, 1974). Researchers of college education even applied different rating instruments to check into various aspects of classroom activities. Over the years, many studies had been undertaken to examine the relationships between the teaching strategies and the measures of the learning outcomes from the students (Brown, 1976; Feldman, 1973; Gessner, 1973; Seldin, 1975).

It seems reasonable to believe that some of the personality variables would play some role in the process of student evaluations. Such a phenomenon is certainly undesirable especially if the ratings are to be considered as valid measures of teaching effectiveness. Several studies were intended to correlate the patterns of student ratings with some of the identified teacher traits like sex, interest, and warmth (Costin & Grush, 1973; Isaacson, McKeachie, and Milholland, 1963; McKeachie & Lin, 1971). Their findings were mixed.

In a recent investigation, Elmore and LaPointe (1975) did not find significant interactions between faculty sex, student sex, and teacher warmth. However, they found out that a different picture was shown when combining both faculty sex and warmth. That is, teacher perceived as warmer or primarily more interested in students tended to receive higher ratings regardless of their sex, whereas female faculty were rated more favorably than their male counterparts when they were low in both traits. It is very clear that personal warmth counts a rather significant weight in determining the rating behaviors of students.

Attempting to investigate the possible relationship between the selected class characteristics and student ratings of instructors, Pohlmann (1975) identified the grades expected by students and the percentage of students taking the course as an elective to be the two most influential in appraising teaching performance. The notable impact of course grades on student evaluations was also found by Brown (1976) and Kennedy (1975) recently. Based on their findings they suggested that higher ratings positively correlated with higher grades the students had earned.

As already identified in many studies that some uncontrolled extraneous variables would confound the results of student ratings, so the constructs of the instruments developed and their reliability still remain open problems. As a matter of fact, no one instrument is proving to be universally acceptable. Factorial structure established from a particular instrument was not easily identified in others. Attributes generated in one study were not obtained in others. Whatever formats they were employed one thing for sure is that the appraisal forms were developed to serve the same function, namely, measuring teaching performance in classroom. Accordingly, two instruments should relate

each other in some respects to some degree when administered to the same groups of students. This is the question to which this study intended to find an answer. In addition, the concurrent validity of the inventories was estimated by correlating the obtained achievement scores with student responses to the two selected appraisal forms.

METHOD

Subjects

For this study, a total of 308 students were sampled from West Chester State College, West Chester, Pennsylvania. With 74% of them being female, the Ss enrolled in 12 classes of undergraduate education courses by 6 instructors. They were almost evenly divided among sophomores, juniors, and seniors. All instructors were male full-time professors of education, had earned the terminal degree, and had at least two years of college teaching experience.

Two evaluation inventories specially developed for higher education level were selected for this study. One was The Inventory of Student Perceptions of Instruction (ISPI) (Scott, 1973): the other used was The Instructional Improvement Questionnaire (IIQ) (Pohlman, 1972). Of the 308 subjects, 211 completed both instruments with additional 54 taking ISPI only and 42 IIQ only.

Procedure

For ascertaining effective evaluation of instruction, it is essential that honest responses by students be solicited and their anonymity warranted. However, these two goals are very difficult, if not completely impossible, to achieve if the instructor is present in the classroom when the student evaluations are underway. Therefore, for this study, a graduate assistant instead of the teaching instructor was asked to conduct the student evaluations.

Furthermore, in order to minimize any possible procedural variations in administration, an instruction sheet with step-by-step directions was prepared for the graduate assistant to use in each of the 12 class.

The ISPI, which consisted of 69 items on a five-point scale, was given about half way through the Fall Semester of 1976 while the IIQ was distributed toward the end of the semester to the same classes. The IIQ contained 40 items divided into two parts on two different sets of five-point scale. The first 20 items were specifically designed for the evaluation of instructor, the next 20 items for course content. Both instruments were administered by the same graduate assistant during the regular class sessions.

At the end of the semester, three separate achievement scores were obtained from each instructor for all the students participated in the study. The first two scores were students performance on the midterms whereas the third was their scores on the final examination. These scores were to be used for further analyses together with their responses on the inventories.

Data analysis

Responses to the two inventories were factor analyzed, separately, through the principal component technique with unities remained in the diagonal elements of the correlation matrix formulated. Factors with eigenvalues greater than one were extracted and rotated using the varimax solution. The generated factor scores were then used as the predictor variables to correlate with the three achievement scores, which here served as the criterion variables. By this means, three different canonical correlation analyses were performed, first the achievement scores with the ISPI factor scores, then, again, the achievement scores with the IIQ factor scores, and finally between the ISPI

and the IIQ factor scores.

RESULTS

The mean and the standard deviation for each of the three achievement scores were presented in Table 1.

Insert Table 1 about here

Six factors were generated with eigenvalues greater than one from the responses of the 266 students to the 69-item ISPI, accounted for nearly half (47%) of the total cumulative variance. As shown in Table 2, the first factor weighed 28.7% of the total variance. For the remaining factors, the variance accounted for was ranging from 5.20 percent to 2.70 percent of the total cumulative variance.

Insert Table 2 about here

For the IIQ, which consisted of 40 items, eight factors with eigenvalues larger than one were extracted and rotated through the varimax method. The eight factors contributed 66.7% to the total variance (see Table 3), in which 37.3 percent was due to the first factor. The weight of the variance accounted for the rest was ranging from 2.6 percent to 7.9 percent.

Insert Table 3 about here

For the three canonical analyses performed in this study, the first analysis, relating the three achievement scores with the six ISPI factor scores generated, produced two highly significant canonical correlation coefficients (see Table 4) as tested against Bartlett's approximate X^2 statistic (Tatsuoka, 1971). The maximum (or the first) canonical correlation coefficient (R_c = .74 with df = 18; p = <.001) indicated the amount of variance shared

by the first two canonical variates was 54% (equivalent to eigenvalue). After removing the first pair of the canonical variates, the second canonical correlation was computed as .41 (with df = 10), which is again significant beyond the .001 level.

Insert Table 4 about here

The results of the second canonical analysis, correlating the achievement scores with the eight factor scores produced from the IIQ, were reported in Table 5. The overall canonical correlation coefficient obtained was significant far beyond the .001 level (R_c = .65 with df = 24; p = <.001), counting off 42% of the variance due to the first pair of canonical variates. A second

---
Insert Table 5 about here
---

significant correlation was computed (R_c = .56 with df = 14; p = <.001) after the first two canonical variates were determined. Also, in this analysis, a third canonical correlation was obtained (R_c = .40 with df = 6; p = .001), which, again, was statistically significant.

To investigate the relationship between these two particular instruments for appraising teaching performance of college professors, a third canonical analysis was carried out by correlating the two sets of the factor scores generated previously. As demonstrated in Table 6, five canonical correlation coefficients were produced, with the first three being significant beyond the .001 level, the fourth one at the .01 level, and the last coefficient at the .05 level, after the first four pairs of canonical variates were removed.

Insert Table 6 about here

DISCUSSION

There were two purposes for which this study was conducted. First, to examine the relationship between college students achievement and their ratings of instructors. Second, to validate the two selected evaluation instruments which were designed specially for assessing teaching performance at higher education level.

From the evidence identified in the separate canonical analyses, the major hypothesis concerning the relationship between the achievement scores and student ratings of teaching effectiveness was supported. This was further confirmed from the fact that the students were asked to evaluate their instructors twice using two different instruments. Their responses to each of the two inventories resulted in consistently significant correlations with the three achievement scores they received at different points of the semester. Their ratings had a great deal of reliability since several precautious steps were taken to maximize the consistency in the administration of the evaluation forms. For example, a graduate assistant rather than the instructor himself conducted the evaluations in the classroom. Furthermore, the prepared written directions instead of the oral instructions were followed when the evaluations were conducted in each of the classes involved. Rather contradictory to what was found in a recent study by White and de la Serna (1976), the findings indicated that college students tend to rate their professors more favorably if they felt they had learned a lot more and achieved much higher in the course. They do not evaluate their professors teaching in an isolated, irrelevant of their own performance.

Additionally, from the highly significant canonical correlation coefficients between the factor scores of the two inventories, an important point

was implied. That is, these two appraisal forms, though not covering exactly the same aspects of classroom instruction, have a great deal of commonality in the measures of teaching effectiveness in college education. Evidently, both had served the intended function well as was attested from the concurrent validity estimated through the multivariate analyses in this study.

Benton, S. E. A comparison of two types of student response inventories for appraising instruction. Unpublished doctoral dissertation, the University of Georgia, Athens, Georgia, 1974.

Brown, D. L. Faculty ratings and student grades: A university-wide multiple regression analysis. Journal of Educational Psychology, 1976, 68, 573-578.

Costin, F., & Grush, J. E. Personality correlates of teacher-student behavior in the college classroom. Journal of Educational Psychology, 1973, 65, 35-44.

Elmore, P. B., & LaPointe, K. A. Effect of teacher sex, student sex, and teacher warmth on the evaluation of college instructors. Journal of Educational Psychology, 1975, 67, 368-374.

Feldman, K. A. Student assessment of teaching: A selected bibliography. Improving College and University Teaching, 1973, 21, 62-63.

Kennedy, W. R. Grades expected and grades received - their relationship to students' evaluations of faculty performance. Journal of Educational Psychology, 1975, 67, 109-115.

Gessner, P. K. Evaluation of instruction. Science, 1973, 180, 566-569.

Isaacson, R. L., McKeachie, W. J., & Milholland, J. E. Correlation of teacher personality variables and student ratings. Journal of Educational Psychology, 1963, 54, 110-117.

McKeachie, W. J., & Lin, Y. G. Sex differences in student response to college teachers: Teacher warmth and teacher sex. American Educational Research Journal, 1971, 8, 221-226.

Pohlmann, J. T. Factor analyses of Parts I and II of the IIQ. Technical Report 6. 1-72, Testing Center, Southern Illinois University, 1972.

Pohlmann, J. T. A multivariate analysis of selected class characteristic and student ratings of instruction. Multivariate Behavioral Research, 1975, 10, 81-92.

Scott, O. The Inventory of Student Perceptions of Instruction: Its history, characteristics and uses. Paper presented at the Twelfth Southeastern Invitational Conference on Measurement in Education, Blacksbury, Virginia, 1973.

Seldin, P. How colleges evaluate professors: Current policies and practices in evaluating classroom teaching performance in liberal arts colleges. Croton-on-Hudson, New York: Blythe-Pennington, 1975.

Tatsuoka, M. M. Multivariate analysis: Techniques for educational and psychological research. New York: John Wiley & Sons, 1971.

White, W. F., & de la Serna, M. Students' ratings of their instructors and the instructors' ratings of students in educational psychology. Journal of Psychology, 1976, 92, 195-198.

TABLE 1

Mean and Standard Deviation of
Achievement Scores

Test	$\bar{X}$	SD	N
First	70.62	19.79	291
Second	73.70	19.68	291
Final	79.83	27.05	290

TABLE 2

Varimax Rotated Factor Matrix of ISPI*
(N = 266)

Item	I	II	III	IV	V	VI**	h^2
1		- 59					.41
2		69					.54
3							.19
4							.46
5		46					.46
6							.52
7	45						.57
8		- 66					.58
9							.43
10				47			.42
11				50			.45
12							.33
13							.45
14				46			.52
15				64			.63
16							.38
17							.29
18				57			.51
19				- 58			.49
20		66					.57
21							.39
22				57			.37
23	66						.52
24	66						.55
25	- 55						.41
26	54						.53
27	59						.49
28	74						.58
29					53		.36
30							.59
31				55			.52
32	51						.58
33					54		.43
34					49		.47
35		46					.78
36							.41
37							.62
38	53						.47
39							.26
40		65					.60
41		69					.65
42		- 53					.49
43						50	.29
44		- 52					.40
45			47				.49
46		56					.50
47	64						.49
48	58						.58
49			- 51				.48
50							.19

Varimax Rotated Factor Matrix of ISPI (con't)

Item	I	II	III	IV	V	VI	h^2
51		47	51				.59
52	57						.61
53			49				.42
54			69				.57
55			67				.61
56			66				.57
57			68				.52
58			- 72				.61
59	62						.51
60							.30
61							.31
62			46				.31
63	57						.47
64							.06
65							.19
66				45			.55
67				54			.40
68	46						.38
69			52				.50
Eigenvalue	19.83	3.56	2.79	2.10	2.01	1.84	
Cum Pct	28.7	33.9	37.9	41.0	43.9	46.6	

---

* Loadings equal to or greater than .45 were entered.

** Decimal points omitted.

TABLE 3

Varimax Rotated Factor Matrix of IIQ*
(N = 254)

Item	I	II	III	IV	V	VI	VII	VIII**	h^2
1						53			.63
2									.54
3			73						.67
4			66						.59
5		73							.65
6		70							.65
7		73							.64
8		65							.71
9									.54
10		68							.61
11	65								.74
12		51							.68
13			75						.70
14			64						.67
15		60							.65
16									.62
17		56							.65
18					63				.66
19		65							.69
20		56							.74
21	86								.82
22	83								.77
23						55			.68
24	- 67								.60
25								55	.57
26								64	.53
27	88								.85
28							69		.59
29	81								.79
30									.49
31	- 87								.82
32									.71
33						58			.54
34	- 72								.69
35					- 68				.54
36				- 65					.69
37	81								.72
38				- 82					.74
39				- 70					.69
40	87								.84
Eigenvalue	14.90	3.19	2.02	1.74	1.46	1.19	1.13	1.05	
Cum Pct	37.3	45.2	50.3	54.6	58.3	61.3	64.1	66.7	

* Loadings equal to or greater than .50 were entered.
** Decimal points omitted.

TABLE 4

Significance Tests of Canonical Correlations Between Achievement Scores and ISPI

No. of Roots Removed	Eigenvalue	Canonical Correlation	Wilk's Lambda	X^2	Df	Significance
0	.54	.74	.36	137.04	18	.000
1	.17	.41	.78	32.71	10	.000

TABLE 5

Significance Tests of Canonical Correlations Between Achievement Scores and IIQ

No. of Roots Removed	Eigenvalue	Canonical Correlation	Wilk's Lambda	X^2	Df	Significance
0	.42	.65	.33	145.72	24	.000
1	.31	.56	.58	73.04	14	.000
2	.16	.40	.84	23.34	6	.001

TABLE 6

Significance Tests of Canonical Correlations Between IIQ and ISPI

No. of Roots Removed	Eigenvalue	Canonical Correlation	Wilk's Lambda	X^2	Df	Significance
0	.90	.95	.04	421.44	48	.000
1	.39	.62	.39	125.59	35	.000
2	.17	.42	.63	61.52	24	.000
3	.13	.36	.76	36.42	15	.002
4	.12	.34	.87	18.41	8	.018